Lower Bounds on Interactive Compressibility by Constant-Depth Circuits

Lower Bounds on Interactive Compressibility by Constant-Depth Circuits

Arkadev Chattopadhyay *
University of Toronto
arkadev@cs.toronto.edu

Rahul Santhanam †
University of Edinburgh
rsanthan@inf.ed.ac.uk

September 4, 2012

Abstract

We formulate a new connection between instance compressibility [HN10]), where the compressor uses circuits from a class C, and correlation with circuits in C. We use this connection to prove the first lower bounds on general probabilistic multi-round instance compression. We show that there is no probabilistic multi-round compression protocol for Parity in which the computationally bounded party uses a non-uniform AC^0-circuit and transmits at most $n/(\log(n))^{\omega(1)}$ bits. This result is tight, and strengthens results of Dubrov and Ishai [DI06]. We also show that a similar lower bound holds for Majority.

We also consider the question of *round separation*, i.e., whether for each $r \geqslant 1$, there are functions which can be compressed better with r rounds of compression than with $r-1$ rounds. We answer this question affirmatively for compression using constant-depth polynomial-size circuits.

Finally, we prove the first non-trivial lower bounds for 1-round compressibility of Parity by polynomial size $\mathrm{ACC}^0[p]$ circuits where p is an odd prime.

*supported partially by an Ontario Ministry of Innovation fellowship and NSERC research grants.
†supported by ESPRC First Grant EP/H05068X/1.

1 Introduction

Consider the following natural communication game between Alice and Bob. Alice is given an input x and she wishes to decide if $x \in L$ for some specified language L. Unfortunately, she only has access to a class C of circuits which are not powerful enough to compute L. However, she is given the option of communicating with Bob, who is trustworthy and computationally unbounded but does not know x. How many bits of information do Alice and Bob need to exchange to decide if $x \in L$? A trivial protocol is for Alice to send x to Bob and Bob to return the answer. Are there problems L for which this is close to the best possible?

We call this game the C-compression game for L. Compression games were defined and studied in the specific case where Alice has the power of polynomial time by Dell and van Melkebeek [DvM10] [1], under the moniker of "oracle communication games". They use a technique of Fortnow and Santhanam [FS11a] to show lower bounds for solving SAT by deterministic multi-round games, under the assumption that the Polynomial Hierarchy does not collapse . In contrast, C is typically a class of non-uniform circuits in our setting, and we are interested mainly in *unconditional* lower bounds. Clearly, such lower bounds can only be shown for C-compression games where there is already a lower bound known for computing Boolean functions, otherwise we cannot even rule out the case that there is a protocol with cost 1.

In this paper, we study C-compression games where C is AC^0 or ACC^0. Dubrov and Ishai [DI06] proved a lower bound which can be interpreted in our setting as saying that there cannot be a 1-round protocol of cost $O(n^{1-\delta})$ for Parity on n bits in the AC^0-compression game, where $\delta > 0$ is any constant. We prove a stronger and more general bound, which applies to *probabilistic* protocols operating in an *arbitrary* number of rounds. We also consider the question of whether r-round protocols are more powerful in general than $r-1$-round protocols, and obtain a separation for each fixed r. Finally, we prove lower bounds for 1-round ACC^0-compression games.

There are several motivations for considering compression games. One natural motivation is to study the trade-off between *communication cost* and *computational complexity*. In a traditional communication complexity setting, each player holds only part of the input, and is unable to solve the problem by itself because of a lack of information. In a traditional complexity theoretic setting, there is only one player (the algorithm), who might find it difficult to solve the problem because of a lack of computational resources. Our setting interpolates between the two. Here, Alice suffers from a computational bottleneck, not having the power to decide $x \in L$ for herself, while Bob suffers from an informational bottleneck, not knowing x. A similar hybrid between computational and informational constraints was studied by Harsha et al. [HIK$^+$07]. However, in their setting, the traditional communication complexity convention of each player having part of the input is maintained. By distinguishing between an informationally-constrained party and a computationally-constrained one, we are able to obtain somewhat cleaner results.

A more immediate motivation comes from the notion of *instance compression*, defined by Harnik and Naor [HN10] and studied in a number of papers since [DI06, FS11b, DvM10]. The traditional notion of solvability of a language L involves obtaining, for each input x, a 1-bit answer indicating whether $x \in L$ or not. A more relaxed notion is to *compress* x, while still preserving information about its membership in L. In other words, the question is whether there is an easily-computable length-decreasing reduction from L to some language L', and if so, how small is the output of the reduction as a function of the input length? Instance compression has a variety of applications in-

[1] Independently, the notion of a compression game was considered by Ishai [Ish11]

cluding cryptography [HN10], reducing the randomness complexity of sampling [DI06], kernelization in parametrized complexity [BDFH08], succinct probabilistically checkable proofs [HN10, FS11b] and completeness of sparse sets [BH08].

Instance compression of length n instances of a language L to length $l(n)$ using circuits from a class C is equivalent to solving the 1-round C-compression game for L with cost $l(n)$. The generalization to multiple rounds is still relevant to the above applications, as well as having particular significance for the study of computationally-bounded leakage resilience by Faust et al. [FRR+10]. Faust et al. show that there is a circuit transformation which converts any circuit into a circuit resilient against leakage functions computable by AC^0 circuits such that the size of the leakage is bounded. This corresponds in a natural way to compression games. Faust et al. prove their result by using the Dubrov-Ishai lower bound for Parity [DI06]. Our results translate to stronger leakage resilience, and for leakage that can occur in multiple rounds so that the total size of the leakage is bounded (corresponding to multi-round compression games).

We next describe our results and techniques in more detail.

1.1 Our Results and Techniques

A natural candidate for lower bounds on AC^0-compression games is Parity, given that we know a lot about how well constant-depth circuits can compute or approximate Parity [Ajt83, FSS84, Hås86]. Dubrov and Ishai [DI06] show, using the method of random restrictions, that for any constant $\delta < 1$, Parity cannot be solved by a 1-round $\mathsf{AC}^0(\text{poly}(n))$-compression game with cost $O(n^{1-\delta})$. Their method does not seem to extend to proving lower bounds close to linear for multi-round protocols or for probabilistic protocols.

We essentially resolve these questions by making a novel connection between probabilistic multi-round C-compression games and *correlation* with circuits in C [2]. We show that any probabilistic multi-round protocol for a C-compression game solving L in which Alice sends at most $c(n)$ bits implies that there is some sequence of circuits in C which have correlation at least $1/O(2^{c(n)})$ with L. Note that the correlation bound depends *only* on the number of bits sent by Alice. Also, the non-uniformity of the circuit class C is crucial in deriving our connection. Using this connection together with recent tight lower bounds on the correlation of Parity with constant-depth circuits due to Impagliazzo, Matthews and Paturi [IMP12], we can show the following tight result:

Theorem 1.1 *The cost of any probabilistic* $\mathsf{AC}^0(\text{poly}(n))$*-compression game solving Parity is* $\Omega(n/(\log(n))^{O(1)})$*. Moreover, this bound is tight in that for any d, Parity can be solved by a deterministic 1-round* $\mathsf{AC}^0(\text{poly}(n))$*-compression game with cost* $O(n/(\log(n))^d)$*.*

Note that the upper bound is for *one-round* protocols while the lower bound is for probabilistic protocols with an arbitrary number of rounds.

Theorem 1.1 has an application to leakage-resilience, and for this application it is important that the lower bound is for Parity. Consider the problem of encoding a secret in the presence of an adversary that may adaptively perform a sequence of measurements on the secret using polynomial-size constant-depth circuits, such that the total number of bits obtained by the adversary is $n/(\log(n))^{\omega(1)}$. Using the proof of Theorem 1.1, it can be shown that the natural XOR-based secret sharing is secure in this setting [Ish11].

[2] It has been pointed out to us by an anonymous referee that a similar connection in a somewhat different setting is implicit in the work of Harsha et al. [HIK+07]

Since our connection between compression and correlation holds generically, we also obtain conditional lower bounds on probabilistic $\mathsf{SIZE}(\text{poly})$-compression for NP based on a plausible complexity assumption. To the best of our knowledge, this is the first evidence that NP is not probabilistically compressible by polynomial-size circuits. Also, using a communication complexity reduction from Parity to Majority, we can show lower bounds for Majority with parameters similar to those in Theorem 1.1.

We next consider the question of round separations for AC^0-compression games. The question of round separations is a classical one in communication complexity introduced by Papadimitriou and Sipser [PS84], and resolved in a sequence of papers [PS84, DGS87, NW93]. The standard example of a problem that is hard for multiple rounds is the Pointer Chasing problem. Unlike in the standard communication setting, in a compression game Alice has the entire input to herself. Hence information-theoretic techniques used in the communication setting cannot be directly implemented in the compression setting. The way we get around this problem is by devising a new Pointer Chasing problem in which Alice has to compute a hard function to determine a pointer. The intuition is that for a computationally bounded Alice, this is the same as a missing pointer and she has to seek Bob's help to determine the right pointer. Justifying this intuition is subtle and requires technical work involving random restrictions. More precisely, we show the following:

Theorem 1.2 *For every constant $r \geq 2$, there exists a Boolean function $T_{r-1}^{m,m}\big(h^{PAR}, Parity\big)$ on $n = O(m^r)$ input bits satisfying the following:*

- *the deterministic $\mathsf{AC}^0\big(\text{poly}(n)\big)$-compression game for the function can be solved with cost $O(m)$ in r rounds.*
- *every probabilistic $\mathsf{AC}^0\big(\text{poly}(n)\big)$-compression game for the function requires cost $\omega\big(m^{2-\epsilon}\big)$ to solve in $r-1$ rounds, for each constant $\epsilon > 0$.*

Finally, we explore the problem of proving incompressibility results for circuit classes for which strong correlation bounds are not known unconditionally. The smallest such natural class of circuits is perhaps AC^0 augmented with MOD_p gates for an odd prime p, known as $\text{ACC}^0[p]$. To the best of our knowledge, no lower bounds were known on even the cost of 1-round compression games. We prove the following:

Theorem 1.3 *Let p be a fixed odd prime. The cost of any 1-round randomized $ACC^0[p]\big(poly(n)\big)$-compression game solving Parity is $\Omega\big(\sqrt{n}/(\log n)^{O(1)}\big)$.*

1.2 Plan of the Paper

In Section 2, we introduce the basic notions needed in this work. In Section 3, we formalize the connection between correlation and compression. We then use it to prove Theorem 1.1 showing the incompressibility of Parity by $\mathsf{AC}^0\big(\text{poly}(n)\big)$ circuits. In Section 4, we establish Theorem 1.2. In Section 5, we give lower bounds for 1-round $\text{ACC}^0[p]$ compression games for Parity, proving Theorem 1.3. Finally, in Section 6, we point out directions for further research.

2 Preliminaries

We assume a basic familiarity with complexity theory. The Complexity Zoo (which can be found at `http://qwiki.caltech.edu/wiki/ComplexityZoo`) is an excellent resource for basic definitions

and statements of results. Another good reference is the book by Arora and Barak [AB09]. We will also be making use of standard concepts from the area of communication complexity [KN97].

We will typically use C to refer to a *class* of (sequences of) circuits in a given format, eg. AC^0 (constant-depth circuits with unbounded fan-in AND and OR gates), $\mathsf{Formula}$ (circuits with binary AND and OR gates), and Ckt (circuits with gates of bounded fan-in). In general, given a circuit class C and a size function $s : \mathbb{N} \to \mathbb{N}$, $\mathsf{C}(s)$ denotes the circuit class C restricted to circuits of size $O(s)$. Occasionally, we will abuse notation and use $\mathsf{C}(s)$ to refer to the class of languages accepted by circuits from C of size $O(s)$. By the size of a circuit, we will always mean the number of wires rather than the number of gates.

We say a class C of circuits is closed under OR if for any sequence of circuit families $\{D_n\}$, where for each n, D_n is a family of circuits from C on n bits, the circuit sequence $\{\vee D_n\}$ belongs to C. For example, the circuit classes AC^0, $\mathsf{Formula}$ and Ckt are closed under OR, but the circuit class AC^0_d of constant-depth circuits of depth d is not. Similarly, we define closure of a class of circuits under AND. A class C of circuits is closed under negation if for any sequence $\{C_n\}$ of circuits, where each C_n is on n bits, the sequence $\{\neg C_n\}$ is also in C. For example, AC^0, $\mathsf{Formula}$ and Ckt are closed under negation but the class of monotone circuits is not.

For a language $L \subseteq \{0,1\}^*$, $L_n = L \cap \{0,1\}^n$.

Given a circuit class C and a language L, the C-compression game for L between two players Alice and Bob is a communication game played as follows. A C-bounded protocol Q in the game consists of a sequence of circuits $\{C_n\}, C_n \in \mathsf{C}$ for Alice and a *strategy* for Bob, i.e., a function from sequences of messages to messages. Initially, Alice has the input x, while Bob has no information. The goal of the game is to decide whether $x \in L$. In a 1-round protocol, Alice sends Bob a single message y_1 obtained by applying $C_{|x|}$ to x, after which Bob announces whether or not $x \in L$. In general, in an r-round protocol, Alice and Bob exchange messages $y_1, z_1, y_2 \dots y_r$, where for each i, y_i is Alice's message in the ith round and z_{i-1} is Bob's message in the i'th round. For each i, y_i is obtained by applying a fixed circuit $C_{|x|}$ to $\langle x, y_1, z_1 \dots z_{i-1} \rangle$, while z_i is an arbitrary function of $y_1, z_1 \dots y_i$, i.e., the history of the protocol so far. We denote the transcript of a protocol Q on input x, i.e., the complete sequence of messages exchanged, by $T_Q(x)$.

A protocol Q *solves* the C-compression game for L if there is a set A such that for each $x \in \{0,1\}^*$, $x \in L$ iff $T_Q(x) \in A$. The communication cost of Q is the total length of messages sent by Alice, i.e., $\Sigma_{i=1}^{r} |y_i|$. Note that we *do not* count the messages sent by Bob when measuring the communication cost. The length of the messages sent by Bob is only restricted implicitly by the fact that Alice uses a circuit $C \in \mathsf{C}$ to compute her messages. If this circuit is polynomial-size, for instance, we can assume wlog that Bob sends only $\mathrm{poly}(n)$ length messages, for any extra message bits cannot affect Alice's messages and hence cannot affect the success of the protocol.

Given functions $c : \mathbb{N} \to \mathbb{N}$ and $r : \mathbb{N} \to \mathbb{N}$, we say that the C-compression game for L can be solved with cost c in r rounds if there is a C-bounded protocol Q solving the game such that on any input of length n, the protocol has cost at most $c(n)$ and uses at most $r(n)$ rounds. We say simply that the C-compression game for L can be solved with cost c if there is a C-bounded protocol solving the game with cost at most c.

Note that for any L and any non-trivial circuit class C, the C-compression game for L can be solved with cost n by a 1-round protocol in which Alice simply sends her input to Bob. Note also that the implicit restriction on the length of Bob's messages via the circuit class C is important - another way of solving a C-compression game is for Bob to send Alice the truth-table of L and Alice to retrieve $L(x)$ from the truth-table.

In case L can be solved by circuits in the class C, the C-compression game has a trivial protocol - Alice decides for herself whether $x \in L$ and sends the answer to Bob. This gives a protocol with cost 1.

As defined above, protocols in the C-compression game are deterministic and solve L on all inputs. We can extend this in a natural way to *probabilistic* and *average-case* C-compression games. In a probabilistic C-compression game, Alice has private randomness and each message of hers is obtained by applying her circuit to the history of the protocol together with her private randomness. A probabilistic protocol Q consists of a sequence of randomized circuits for Alice and a strategy for Bob. For error function $\epsilon : \mathbb{N} \to [0,1]$ and a cost function $c : \mathbb{N} \to \mathbb{N}$, the protocol solves L with cost c and error at most ϵ if the total length of messages sent by Alice on any run of the protocol is at most $c(|x|)$ and there is a set A such that if $x \in L$, then $\Pr(T_Q(x) \in A) \geqslant 1 - \epsilon(|x|)$, and if $x \notin L$, then $\Pr(T_Q(x) \in A) \leqslant \epsilon(|x|)$. Given a function $q : \mathbb{N} \to [0,1]$, an average-case protocol Q for L with success rate q is a deterministic protocol such that there is a set A for which, for at least $q(n)$ fraction of inputs x of length n, $x \in L$ iff $T_Q(x) \in A$. If the circuit class C is non-uniform, then any probabilistic protocol with error at most $\epsilon(n)$ can be converted to an average-case protocol with success rate $1-\epsilon(n)$ simply by fixing the private randomness of Alice so as to maximize the success rate.

We seek to prove upper and lower bounds on the cost of compression games for interesting languages L and classes C of circuits. For most of the paper, the focus will be on AC^0, the class of polynomial-size constant-depth circuits with AND and OR gates, where the gates have unbounded fan-in. We always measure size as a function of the input length $|x|$.

One of the main ideas in our paper is to connect cost of C-compression games with correlation bounds against C. Given a class C of circuits, a language L and a function $s : \mathbb{N} \to [0,1]$, L has *correlation* at most s with C if for any circuit $C \in \mathsf{C}$ and all $n \in Nat$, $\Pr_{x \in \{0,1\}^n} C(x) = L(x) \leqslant 1/2 + s(|x|)/2$.

The following inequality, called the Chernoff bound, will be useful in Section 4. We denote the expectation of a random variable X by $\mathbb{E}[X]$.

Theorem 2.1 (Chernoff bound) *[DP09] Let $X = \sum_i X_i$ be a sum of independent random variables, each of which takes value in $[0,1]$. Then, $\Pr\big[|X - \mathbb{E}[X]| > \epsilon \cdot \mathbb{E}[X]\big] < 2 \cdot \exp\big(-\epsilon^2/3 \cdot \mathbb{E}[X]\big)$, where $\epsilon > 0$ is any constant.*

3 Compression implies correlation

In this section, we show that for classes of circuits C closed under OR and negation, if the C-compression game for L can be solved with low cost, then L correlates well with some circuit in C. We show this first for deterministic compression games, and then extend the argument to probabilistic and average-case games. A crucial feature of our connection between compression and correlation is that it works for multi-round games - this enables us to strengthen and generalize the lower bound of Dubrov and Ishai [DI06] for solving Parity with 1-round AC^0-compression games.

First, we require the following folklore lemma saying that if a language is computed by an OR of circuits from a class C which is not too large, then it correlates reasonably well with some circuit in C. This lemma follows, for instance, from the Discriminator Lemma [HMP$^+$93].

Lemma 3.1 *Let C be any circuit class containing circuits for the constant functions 0 and 1. Let $f : \mathbb{N} \to \mathbb{N}$ be a function such that $f(n) \geqslant 2$ for all n, and $L \subseteq \{0,1\}^*$ be a language such that for*

each n, L_n is computed by the OR of at most $f(n)$ circuits from C. Then L has correlation at least $1/O(f(n))$ with C.

Proof: Fix n and let $C_{i,n}, 1 \leqslant i \leqslant f(n)$ be a family of $f(n)$ circuits each on n bits from the class C such that L_n is the OR of some subset of those circuits. If smaller than a $1/2-1/(2f(n))$ fraction of strings of length n belong to L_n, then L_n has correlation at least $1/f(n)$ with the constant function 1 and hence with C. So assume that at least a $1/2-1/(2f(n))$ fraction of strings of length n belong to L_n. Then, since L_n is computed by the OR of the $C_{i,n}$'s, there must be some j such that $C_{j,n}$ is 1 for at least a $1/(4f(n))$ fraction of strings of length n; moreover each 1-input to $C_{j,n}$ is a 1-input to L_n. Consider the set of inputs X_n of length n for which $C_{j,n}$ evaluates to 0. If L_n is 0 for at least half these inputs, then $C_{j,n}$ has correlation at least $1/(4f(n))$ with L_n, otherwise the constant function 1 has correlation at least $1/(4f(n))$ with L_n. In either case, L_n has correlation at least $1/(4f(n))$ with C. □

Lemma 3.2 *Let $c : \mathbb{N} \to \mathbb{N}$ be a function such that $c(n) \leqslant n$ for all n, C be a class of circuits closed under OR and negation, $s : \mathbb{N} \to \mathbb{N}$ be a size function such that $s = \Omega(n)$, and L be a language. If there is a $\mathsf{C}(s(n))$-compression game for L with cost at most $c(n)$, then L has correlation at least $1/O(2^{c(n)})$ with $\mathsf{C}(s(n))$.*

Proof: Suppose there is a $\mathsf{C}(s(n))$-compression game for L with cost at most $c(n)$. Let $\{C_n\}$ be the sequence of C-circuits used by Alice in her protocol, with the size of each circuit C_n being at most $s(n)$, and let f be Bob's strategy. We define some notions that will be useful in the proof.

A candidate transcript $T =< y_1, z_1, y_2 \ldots y_r >$ is simply a tuple of strings which can be interpreted as a sequence of messages in the protocol. Note that a candidate transcript might not actually correspond to any real protocol. We say that a candidate transcript is *Bob-consistent* if for each $i, 1 \leqslant i \leqslant r-1$, $z_i = f(y_1 \ldots y_i)$. Informally, a Bob-consistent candidate transcript looks OK from Bob's point of view, in that every message z_i is actually obtained by applying his strategy f to the history so far. A simple, but crucial, point is that the question of whether a candidate transcript is Bob-consistent depends only on the transcript itself, and not on x. This is because Bob has no information about x - his view of the protocol is defined entirely by messages from Alice.

We say that a candidate transcript is Alice-consistent on an input x if for each $i, 1 \leqslant i \leqslant r$, $y_i = C_{|x|}(x, y_1, z_1 \ldots z_{i-1})$. Namely, Alice's message is actually obtained by applying the appropriate circuits $C_{|x|}$ to the history so far. We say that a candidate transcript is consistent on input x if it is both Bob-consistent and Alice-consistent on x. Moreover, we say that a candidate transcript is accepting if after receiving the message y_r, Bob announces that the input is in L. Note that again the question of whether a transcript is accepting depends only on the transcript and not on x. We say that the candidate transcript is t-bounded if $\sum_{i=1}^{r} |y_i| \leqslant t$.

Now, $x \in L$ iff there is a candidate transcript $T =< y_1, z_1 \ldots y_r >$ such that T is consistent on x and accepting, and moreover T is $c(|x|)$-bounded. One direction of this claim is immediate - if $x \in L$, then the transcript of the protocol given by $C_{|x|}$ and strategy f for Bob is consistent and accepting, and satisfies the condition that the total length of messages sent by Alice is at most $c(|x|)$. Conversely, suppose there is a candidate transcript T that is consistent on x and accepting. Since the protocol is Alice-consistent on x, we have that y_1 is indeed the first message sent by Alice. Since the protocol is Bob-consistent, we have that z_1 is indeed the first message sent in response to Bob. Continuing inductively, we have that for each round i, the messages sent by Bob and Alice

are indeed z_{i-1} and y_i. Since the transcript is accepting, we have that Bob does accept at the end of the protocol, which implies $x \in L$ by the assumption that the protocol is a correct protocol for the C-compression game for L.

We would like to take advantage of this characterisation to design circuits checking if $x \in L$. The idea is to cycle over Bob-consistent accepting $c(|x|)$-bounded candidate transcripts checking for each one whether it is Alice-consistent or not. Doing this exhaustively would take exponential size, but in fact we can write the global check as an OR of small circuits, where the OR is not too large. This will imply that L correlates reasonably well with some circuit in C, by using Lemma 3.1.

Now consider any Bob-consistent accepting $c(|x|)$-bounded candidate transcript T. Note that there are at most $2^{c(|x|)}$ such transcripts, even though we are placing no a priori bound on the length of Bob's messages. For each sequence of messages $y_1, y_2 \dots y_r$ sent by Alice of total length at most $c(|x|)$, since Bob's strategy is deterministic, there is at most one Bob-consistent accepting candidate transcript containing these messages in the y-positions of the tuple.

For each T as described in the paragraph above, we construct a circuit C'_T which checks whether T is Alice-consistent. The key idea here is *local checkability* - rather than simulating a run of the protocol, C'_T checks in parallel for each round whether the message sent by Alice in that round is consistent with the history. Thus the top gate of C'_T is an AND gate of fan-in r, where r is half the number of elements in the tuple T. The i'th input to the AND gate, $1 \leqslant i \leqslant r$, is a circuit checking whether y_i is consistent with $x, y_1 \dots z_{i-1}$. This is done simply by simulating $C_{|x|}$ on $< x, y_1 \dots z_{i-1} >$ and checking using $O(|y_i|)$ OR and negation gates whether the output is precisely y_i.

For each T which is Bob-consistent, accepting and $c(|x|)$-bounded, the total size of C'_T is at most $r + \sum_{i=1}^{r} |y_i| + O(s)$. This is $O(s)$ since $c(n) \leqslant n$ and $s(n) = \Omega(n)$. By the assumption that the circuit class C is closed under OR and negation, we have that each circuit C'_T belongs to C, moreover it is in $\mathsf{C}(s)$ by the previous line.

Now, by the characterization of L in terms of consistent accepting $c(n)$-bounded transcripts, we have that for each n, L_n is computed by the OR over the at most $2^{c(n)}$ Bob-consistent accepting candidate transcripts T of C'_T. Applying Lemma 3.1, this implies that L has at least correlation $1/O(2^{c(n)})$ with $\mathsf{C}(s)$. □

We apply Lemma 3.2 to obtain lower bounds on AC^0-compression for the Parity language. Dubrov and Ishai [DI06] considered this question. In our terminology, they study the cost of 1-round AC^0-compression games for Parity. They showed that for any constant $\delta > 0$, Parity cannot be solved with a 1-round compression game of cost $n^{1-\delta}$. We improve this bound, and more significantly, extend it to the setting of multi-round games. To this end, we exploit the connection with correlation given by Lemma 3.2, and use the following recent result of Impagliazzo, Mathews and Paturi [IMP12], which settles an open problem posed by Hastad in his doctoral dissertation [Hås87] [3].

Theorem 3.3 *[IMP12] For any size function $s : \mathbb{N} \to \mathbb{N}$ and positive integer d, Parity has correlation at most $2^{-n/O((\log(s))^{d-1})}$ with AC^0-circuits of size s and depth d.*

[3] Independently, and around the same time, Hastad himself has proved a version of the following result with slightly weaker parameters using a somewhat different technique. His result is still unpublished.

Theorem 3.4 *The cost of any* $\mathsf{AC}^0(\text{poly}(n))$*-compression game solving Parity is* $\Omega(n/(\log(n))^{O(1)})$*. Moreover, this bound is tight in that for any d, Parity can be solved by a 1-round* $\mathsf{AC}^0(\text{poly}(n))$*-compression game with cost* $O(n/(\log(n))^d)$*.*

Proof: Suppose there is an $\mathsf{AC}^0(\text{poly}(n))$-compression game solving Parity with cost $c(n)$. By Lemma 3.2, Parity has correlation at least $1/O(2^{c(n)})$ with polynomial-sized AC^0-circuits of depth d, for some fixed d. By Theorem 3.3, Parity has correlation at most $2^{-n/O(\log(n))^{d-1}}$ with AC^0-circuits of $\text{poly}(n)$ size and depth d. Thus we get that $c(n) = \Omega(n/(\log(n))^{O(1)})$.

To show that the bound is tight, we use the fact that for any d, Parity can be solved on instances of length $(\log(n))^d$ by polynomial-sized AC^0-cricuits of depth $d+1$ just by a simple divide-and-conquer technique. This gives the following strategy for Alice in a 1-round AC^0-compression game for Parity. She divides the input into $n/(\log(n))^d$ blocks of $\log(n)^d$ bits each (we assume for simplicity here that n is a power of two - this doesn't affect the asymptotics). She computes Parity on each block using polynomial-sized AC^0 circuits of depth d and sends the resulting values to Bob. Bob computes the parity of the bits he sent and accepts iff the computed value is 1. The cost of this protocol is $O(n/(\log(n))^d)$. □

Apart from the fact that Theorem 3.4 says something interesting about games with an arbitrary number of rounds, one advantage of the proof technique is that complexity lower bounds for circuit classes yield communication lower bounds for the compression game in a modular fashion. In contrast, the proof of Dubrov and Ishai [DI06] adapts the classical random restriction technique used to prove constant-depth circuit lower bounds to the setting of compression.

Perhaps the biggest advantage of our proof technique, though, is that it says something about *probabilistic compression*. In the setting of parameterized instance compression [FS11a, DvM10], getting complexity-theoretic evidence against general probabilistic compression of NP problems is a major open question. In our setting of AC^0-compression games, we are able to resolve this question for the Parity problem, and indeed for any language which has small correlation with constant-depth circuits.

Our lower bounds work in the more general setting of average-case compression. The natural strategy is to prove an analogue of Lemma 3.1 saying that if an OR of circuits correlates well with some Boolean function, then one of the circuits correlates well with the function. Unfortunately, this is not true in general. Instead, we show a refined version stating that if an OR of *disjoint* circuits (namely, circuits such that no two different ones output 1 on the same input) correlates well with some *balanced* Boolean function, then one of the circuits correlates well with the function. Then, taking advantage of the structure of the Proof of Lemma 3.2, we are able to establish a connection between average-case compression and correlation.

The following lemma is new to the best of our knowledge, and might be of independent interest.

Lemma 3.5 *Let* C *be any circuit class. Let* $f : \mathbb{N} \to \mathbb{N}$ *be any function, and* $\{F_n\}$ *be a sequence of families of circuits from* C *such that for each n,* F_n *contains at most* $f(n)$ *circuits, each one on n bits, satisfying the condition that for each input* $y \in \{0,1\}^n$ *and distinct circuits* $C_1, C_2 \in F_n$*, either* $C_1(y) = 0$ *or* $C_2(y) = 0$*. Let* $\epsilon : \mathbb{N} \to [0,1]$ *be an arbitrary function, and* $L \subseteq \{0,1\}^*$ *be a balanced language (i.e.,* L_n *has exactly* 2^{n-1} *strings for each n) such that for each n,* L_n *has correlation at least* $\epsilon(n)$ *with the OR of circuits in* $\{F_n\}$*. Then there is a sequence of circuits* $\{C_n\}$ *such that for each n,* $C_n \in F_n$ *and* C_n *has correlation at least* $\epsilon(n)/f(n)$ *with* L_n*.*

Proof: Fix n and let the circuits in F_n be $C_1, C_2 \dots C_k$, where $k \leqslant f(n)$. By assumption, the circuits in F_n all have disjoint 1-sets, and the OR of the circuits, denoted by C, has correlation at least $\epsilon(n)$ with L_n. We think of L_n as a Boolean function f. It would be convenient to assume that each C_i, C and f outputs a value in $\{1, -1\}$ (with 0 mapped to 1 and 1 to -1).

In this setting, we make the following two observations. First, the correlation between any C_i and f is just $\left|\mathbb{E}_x[f(x)C_i(x)]\right|$. Second,

$$C(x) = \sum_{i=1}^{k} C_i(x) - (k-1)$$

Hence, by linearity of expectation and using the fact that f is balanced, we get

$$\epsilon(n) \geq \left|\mathbb{E}_x\big[C(x)f(x)\big]\right| = \left|\sum_{i=1}^{k} \mathbb{E}_x\big[C_i(x)f(x)\big]\right|$$

By triangle inequality and averaging, there exists an i such that $\left|\mathbb{E}_x[C_i(x)f(x)]\right| \geq \epsilon(n)/k$, which finishes the argument. □

Lemma 3.6 (Compression-Correlation) *Let $c : \mathbb{N} \to \mathbb{N}$ be a function such that $c(n) \leqslant n$ for all n, C be a class of circuits closed under negation, $s : \mathbb{N} \to \mathbb{N}$ be a size function such that $s = \Omega(n)$, and L be a balanced language. Let $q : \mathbb{N} \to [0, 1]$ be a function such that $q(n) \geqslant 1/2$ for all n. If there is an average-case $\mathsf{C}(s(n))$-compression game for L with cost at most $c(n)$ and success rate at least $q(n)$, then there exists circuits $C_1, \dots, C_{c(n)}$, each $C_i \in \mathsf{C}(s(n))$, such that L has correlation at least $(2q(n) - 1)/O(2^{c(n)})$ with the circuit* $\text{AND} \circ \big(C_1, \dots, C_{c(n)}\big)$.

Proof: The proof follows the lines of the proof of Lemma 3.2. The main observation is that the circuits C'_T are disjoint, and hence we can apply Lemma 3.5. The OR of the circuits C'_T will have correlation at least $2q(n) - 1$ with L_n by the assumption on the success rate of the average-case protocol, and hence we get that $\mathsf{C}(s(n))$ has correlation at least $(2q(n) - 1)/O(2^{c(n)})$ with L. □

In the next subsections, we exploit the Compression-Correlation Lemma to show that Parity remains strongly incompressible by some natural classes of bounded depth circuits.

3.1 Application to AC^0

The following is the first strong lower bound on probabilistic multiround compression by a very natural and well studied class of circuits. It significantly extends the earlier lower bound for 1-round compression obtained by Dubrov and Ishai [DI06].

Theorem 3.7 *The cost of any probabilistic $\mathsf{AC}^0(\text{poly}(n))$-compression game solving Parity with error $1/2 - 1/2^{n^{o(1)}}$ is $\Omega(n/(\log(n))^{O(1)})$.*

Proof: Using Lemma 3.6 and Theorem 3.3, we have that any average-case $\mathsf{AC}^0(\text{poly}(n))$-compression game solving Parity with success rate $1/2+1/2^{n^{o(1)}}$ has cost $\Omega(n/(\log(n))^{O(1)})$. The theorem follows from this and the fact that any probabilistic protocol with error at most $\epsilon(n)$ yields an average-case protocol with success rate at least $1 - \epsilon(n)$. □

3.2 Circuits with only MOD_p gates

Finally, we show incompressibility by bounded depth circuits comprising only MOD_p gates, when p is a fixed prime. In order to do this, we will make use of the following correlation bound that is implicit in the work of Chattopadhyay and Wigderson [CW09].

Theorem 3.8 *Let $C_1, \ldots, C_t$ be any circuits of depth d of any size on n input variables, comprising only MOD_p and NOT gates where p is an odd prime. Let f be the boolean function computed by the circuit $\mathrm{AND} \circ \big(C_1, \ldots, C_t\big)$. Then,*

$$\mathrm{Corr}\big(f, \mathrm{PARITY}\big) \leq \exp\Big(- \alpha(p)^{(p-1)^d} \cdot n \Big),$$

where $\alpha(p)$ is a constant determined by p.

Combining our Compression-Correlation Lemma with Theorem 3.8, we immediately get the following very strong bound:

Theorem 3.9 *Let p be any fixed odd prime. The cost of any probabilistic $\mathrm{CC}^0[p](s(n))$-compression game solving Parity with error $1/2 - 1/2^{o(n)}$ is $\Omega(n)$.*

3.3 More general circuits

Since the Compression-Correlation Lemma is very general, it can be used to derive compression lower bounds for C-compression games for larger classes C under complexity assumptions. As an example, we have the following result:

Corollary 3.10 *Suppose there is a language $L \in \mathsf{NP}$ such that any sequence of polynomial-size circuits has correlation at most $1/2^{n^{\Omega(1)}}$ with L. Then the $\mathsf{SIZE}(\mathrm{poly}(n))$-compression game for L has cost $\Omega(n^{\Omega(1)})$.*

As far as we are aware, this is the first lower bound on probabilistic multi-round (or even single-round) compression for NP based on a plausible complexity assumption relating to solvability by polynomial-size circuits.

A natural question is whether our techniques can be applied to get lower bounds in the C-compression game for some Boolean function f for which it is *known* that f correlates well with C. The answer is positive: we are able to show similar lower bounds as in Theorem 3.4 for solving the Majority problem using AC^0-compression games. The Majority problem asks whether at least half the bits in the input are 1. Note that Majority is a monotone function, and that any monotone function is known to have correlation at least $\log(n)/n$ with one of its input bits by a classic result of Kahn, Kalai and Linial [KKL88].

The key idea in showing the lower bound for Majority is to reduce from Parity to Majority within the setting of compression games. We only know how to do the reduction using a multi-round compression game where the number of rounds grows with n, but here we reap the advantages of proving a lower bound for Parity in AC^0-compression games with an *arbitrary* number of rounds.

Lemma 3.11 *Let $c : \mathbb{N} \to \mathbb{N}, c(n) \leqslant n$ and $r : \mathbb{N} \to \mathbb{N}, r(n) \leqslant n$ be functions. Suppose that the $\mathsf{AC}^0(\mathrm{poly}(n))$-compression game for Majority can be solved with cost $c(n)$ in $r(n)$ rounds. Then the $\mathsf{AC}^0(\mathrm{poly}(n))$-compression game for Parity can be solved with cost $c(2n)\lceil\log(n)\rceil$ in $r(2n)\lceil\log(n)\rceil$ rounds.*

Proof: Suppose that there is an efficient protocol to solve Majority in the AC^0-compression game. The idea is to use this protocol together with binary search to find the weight $wt(x)$ of the input x. By weight here, we mean the number of 1's in the input. Given the weight of the input x, Parity can be decided easily by a low-cost 1-round protocol.

Let x of length n be the input to Alice. We assume wlog that n is a power of 2 - if not, Alice simply pads her input to the next highest power of 2, adding equal numbers of zeroes and ones if n is even and one more zero than ones if n is odd. We define the following protocol between Alice and Bob, which is divided into "super-rounds". Each super-round consists of an execution of an efficient protocol Q for Majority on some input. In the first super-round, Alice and Bob solve the question of whether the majority of bits in Alice's input x are 1 or not. If the answer is "yes", Alice pads her input with $n/2$ 0's and runs the protocol Q on the new input. Namely, she tries to determine whether $wt(x) \geqslant 3n/4$ or not. If the answer is "no", Alice pads her input with $n/2$ 1's and runs the protocol Q on the new input. Namely, she tries to determine whether $wt(x) \geqslant n/4$ or not. In general, after the i'th super-round, Alice knows that the weight of x lies in some interval of length at most $n/2^i$, and in the next super-round, she tries to reduce the size of this interval by a factor of 2.

After $\log(n)$ many super-rounds, Alice knows the weight of the input, which implies that she also knows the parity (since a constant-depth circuit of polynomial size can decide parity given the weight of the input). The total number of rounds of the protocol is at most $r(2n)\lceil\log(n)\rceil$, since the protocol Q is run at most $\lceil\log(n)\rceil$ times, each time on an input of size at most $2n$. The total cost is at most $c(2n)\lceil\log(n)\rceil$, again by the same argument. The circuits used by Alice to implement this reduction are still constant-depth polynomial size. This completes the proof. □

Theorem 3.12 *The* $\mathsf{AC}^0(\mathrm{poly}(n))$*-compression game for Majority cannot be solved with cost* $O(n/(\log(n))^{O(1)})$.

Proof: Suppose the $\mathsf{AC}^0(\mathrm{poly}(n))$-compression game for Majority can be solved with the stated cost. Then, by applying Lemma 3.11, we get that the $\mathsf{AC}^0(\mathrm{poly}(n))$-compression game for Parity can be solved with cost $O(n/(\log(n))^{O(1)})$, which contradicts Theorem 3.4. □

4 The power of interaction

In order to separate the power of $r+1$ round compression from r round compression, we introduce the notion of a tree function, inspired by pointer chasing problems defined in standard 2-party communication complexity [PS84, DGS87, NW93]. Fix a *pointer* function $h : \{0,1\}^m \to [\ell]$ and a Boolean function $f : \{0,1\}^m \to \{0,1\}$. Then, for each integer $i \geq 1$, we define the boolean tree function $TF_i^{m,\ell}(h,f)$ of height i composing h and f as follows. Let T_i^ℓ denote the complete ℓ-ary tree of height i. For $i \geq 1$, the input of $TF_i^{m,\ell}(h,f)$ is a boolean string of length $m(1+\ell+\cdots+\ell^{i-1})$ that is interpreted to assign each node of tree T_i^ℓ with an m-bit label in the following natural way: the first m bits of the input label the root of T_i^ℓ. The next $m\ell$ bits of the input are grouped into ℓ equal sized blocks $C_1,\ldots,C_\ell$, where each block C_i has m bits. Each of the m blocks is used to label a distinct node at level 1 of the tree. Proceeding in this way, we assign labels to all nodes of the tree T_i^ℓ. We define the tree function $TF_i^{m,\ell}(h,f)$ by induction on i: $TF_1^{m,\ell}(h,f)$ first evaluates h on the label y of the root node of the tree T_1^ℓ to obtain the index of a child of the root. Then f is applied to the label of the pointed child node. In general, $TF_i^{m,\ell}(h,f)$ for $i > 1$, evaluates h

on the label of the root node of T_i^ℓ to travel to a child node Q. Then, we apply $TF_{i-1}^{m,\ell}(h,f)$ to the string formed by concatenating the labels of the nodes of the subtree of height $i-1$ rooted at Q.

Note that for any reasonably powerful circuit class C, such as AC^0, and for any integer $r \geqslant 2$, there is a simple deterministic protocol solving the C-compression game for $TF_{r-1}^{m,m}\big(h,f\big)$ in r rounds with cost $O(m)$. This is because, starting with the root node label, Alice can send Bob the label of the current node. Bob responds by evaluating h on it, expecting Alice to send back the label of the relevant child. The interaction continues until Alice sends the label of the relevant leaf, at which point Bob evaluates f on the label and thereby decides whether the input is a YES input.

The question we want to understand is what happens when the game has only $r-1$ rounds. It would seem that if h is extremely hard for C and f is incompressible by C, then the best that a protocol with $r-1$ rounds can do is follow the r round game until the $r-2$th round, and then in the final round Alice transmits the m-bit label of the relevant node at level $r-2$ in T_{r-1}^m along with the labels of all its m children. In such a protocol, in the final round, Alice communicates $\Omega(m^2)$ bits. A natural question is to understand for which h, f and C this is unavoidable in C-compression games. In this section, we describe a simple h and f for which the above is essentially an optimal strategy to follow for $\mathsf{AC}^0\big(\mathrm{poly}(n)\big)$-compression games.

We will use Parity as our function f, and the pointer function h is also based on Parity as follows. Divide the m bits of input to the pointer function into $\log(\ell)$ equal sized blocks. We assume wlog that ℓ is a power of 2, and that $\log(\ell)$ divides m. First, h^{PAR} evaluates the parity of each block to generate a $\log(\ell)$ bit string y. Then, it outputs the number in $[\ell]$ whose binary encoding is y. We now state the main theorem of this section.

Theorem 4.1 (restatement of Theorem 1.2) *For every constant $r \geq 2$, the function $TF_{r-1}^{m,m}\big(h^{PAR}, Parity\big)$ on $n = O(m^r)$ input bits satisfies the following:*

- *there is a deterministic r-round $\mathsf{AC}^0\big(\mathrm{poly}(n)\big)$-compression game of cost $O(m)$ that solves it.*
- *every $(r-1)$-round probabilistic $\mathsf{AC}^0\big(\mathrm{poly}(n)\big)$-compression game solving it has cost $\omega\big(m^{2-\epsilon}\big)$, for each constant $\epsilon > 0$.*

Our argument for proving this theorem is based on the random restriction method. The random restriction method was developed in a series of works by [Ajt83, FSS84, Yao86, Cai86, Hås87] for showing that Parity is not efficiently approximable by AC^0 circuits. Below, we briefly recall the relevant results needed for our argument.

A *restriction* ρ is a map from the set of input indices $[n] \to \{0,1,*\}$. The indices that are mapped to 0/1 by ρ are called fixed and the ones that are assigned $*$ are called free. For any function f on $\{0,1\}^n$ and any restriction ρ to its variables, we denote by f_ρ the boolean function naturally induced from f on variables left free by ρ. The reason why restrictions are going to be useful for us is that f_ρ becomes simple for a "typical" ρ. More precisely, let $0 \leq p \leq 1$ be any real number. Then, let Π_n^p be the distribution on restrictions generated by independently fixing each input variable with probability $(1-p)/2$ (respectively $(1-p)/2$) to 0 (respectively 1) and leaving it free with probability p.

Lemma 4.2 *[CW04] Let f be a CNF (or DNF) formula with clause width t on m variables. Let ρ be a random restriction sampled from Π_n^p. Then, there exists a constant $\gamma > 0$ such that the probability of f_ρ not having a decision tree of height at most s is less than $\big(\gamma pt\big)^s$.*

A well known consequence of Lemma 4.2 is the following corollary, whose proof we furnish below for the sake of completeness.

Corollary 4.3 *Let f be a function computed by an AC^0 circuit of size S and depth d and $c > 0$ be some constant. Let $p = 1/\big((2\gamma c)^d (\log n)^{d-1}\big)$. Then*

$$\Pr_{\rho \sim \Pi_n^p}\left[h\big(f_\rho\big) \; > \; c \log n\right] \leq S \cdot \frac{1}{n^c}$$

where $h\big(f_\rho\big)$ denotes the height of the best decision tree for f_ρ.

Proof: This can be shown by a simple inductive argument using Lemma 4.2. Assume, as our inductive hypothesis, the following: let $i \geq 2$ and $p_i = 1/\big((2\gamma c)^i (\log n)^{i-1}\big)$. Let G_i be the set of gates in the ith layer of C and let S_i be the number. Further, let $S_{\leq i} = \sum_{j=1}^{i} S_j$. Our inductive hypothesis is the following:

$$\Pr_{\rho \sim \Pi_n^{p_i}}\left[\exists g \in G_i \; : \; h\big(f_\rho^g\big) \; > \; c \log n\right] \leq S_{\leq i} \cdot \frac{1}{n^c},$$

where f^g is the function computed at gate g. Now, if the ith layer of the circuit has AND (OR) gates then one can assume w.l.o.g that $i+1$th layer has OR (AND) gates. In this case, assuming that each f_ρ^g has a decision tree of height at most $c \log n$, we represent f_ρ^g as a DNF of width at most $c \log n$ by using the small height decision tree. This collapses layers i and $i+1$ and hence the output of every gate at layer $i+1$ is a DNF of width $c \log n$ under the restriction ρ. We apply Lemma 4.2 to each such DNF by hitting with a fresh random restriction of the free variables with $p = 1/\big(2\gamma c \log n\big)$ and $t = c \log n$. Clearly, the probability that any fixed such DNF under the next round of restriction fails to have a decision tree of height at most $c \log n$ is at most $2^{-c \log n} = n^{-c}$. Applying the union bound to S_{i+1} such DNF's (one for each gate at layer $i+1$) immediately completes the induction, modulo establishing the base case of $i = 1$. When $i = 1$, each gate is a CNF or DNF of width 1. Thus, using a restriction with $p = 1/(2\gamma)$ in the Switching Lemma easily establishes the desired base case. □

Corollary 4.3 and Lemma 4.2 will be our main tool for simplifying the circuits employed in an AC^0-compression game. The whole point will be to argue that while the circuits will simplify, the pointer function still remains complex over the free variables.

Theorem 4.1 easily follows from the main technical result of this section stated below, by a standard application of Yao's minmax principle.

Theorem 4.4 *Let $\delta < 1$ be any fixed number and $r \geq 2$ be an arbitrary integer. Then, every $(r-1)$-round deterministic $\mathsf{AC}^0\big(\mathrm{poly}(n)\big)$-compression game of cost $\omega\big(m^\delta \ell\big)$ solves $TF_{r-1}^{m,\ell}\big(h^{\mathrm{PAR}}, Parity\big)$ correctly for at most $1/2 + o(1)$ fraction of the inputs, if $m = \ell^{\Theta(1)}$.*

4.1 Warm-up: 2 vs 1 rounds

We first present the separation of the power of 2-round compression games from that of 1-round compression games. In addition to this being a clean way of conveying our main idea, it also is the base case of our inductive argument for Theorem 4.4.

Lemma 4.5 *Let $\delta < 1$ be any constant. Every 1-round deterministic $\mathsf{AC}^0\big(\mathrm{poly}(n)\big)$-compression game solving $TF_1^{m,\ell}\big(h^{\mathrm{PAR}}, Parity\big)$ with cost $O\big(m^\delta \ell\big)$ errs on at least $1/2 - o(1)$ fraction of the inputs, if $m = \ell^{\Theta(1)}$.*

Proof: Let $C = C_1, \ldots, C_t$ be a circuit used by Alice to send t bits to Bob in a 1-round game, where the ith bit is the output of C_i. Let each C_i be a circuit of size at most n^c and depth d, where c, d are arbitrary constants. We show below that if $t = O\big(m^\delta \ell\big)$ for any $\delta < 1$, the protocol errs on at least $1/2 - o(1)$ fraction of the inputs.

Assume $t = m^\delta \ell$ for some $\delta = 1 - \beta_0$ with $\beta_0 > 0$ being a constant. Let $\beta < \beta_0$ and $c' = c + 2$ be new constants. We will first show that if we sample a restriction ρ from Π_n^p, with $p = 1/\big(2m^\beta (2\gamma c')^d (\log n)^{d-1}\big)$, then each of the circuit C_i simplifies a lot. In particular, each restricted C_i depends on a constant number of free variables with high probability. From this we will conclude that the labels of almost all leaf nodes of T_1^m have many free variables and yet C^ρ depends on only a few free variables from each such label. It will be simple to derive our claim from this. Forthwith the details.

It will be convenient to sample ρ by first sampling ν from $\Pi_n^{p_1}$ with $p_1 = 1/\big(2(2\gamma c')^d (\log n)^{d-1}\big)$ and then sampling η from $\Pi_{n'}^{p_2}$, where n' is the number of free variables in ν and $p_2 = 1/m^\beta$. Invoking Corollary 4.3 and union bound, with probability $1 - O(t/n^{c'-c})$, each C_i^ν has a decision tree of height less than $(c'+1)\log n$. Hence, each C_i^ν is now expressible as a CNF of width less than $(c'+1)\log n$. Let n' be the number of free variables. We hit C^ν with a second random restriction η sampled from $\Pi_{n'}^{p_2}$. Let w be any positive integer. Then, using the Switching Lemma for CNF's of width less than $(c'+1)\log n$, some $C_i^{\eta\nu}$ does not have a decision tree of height less than w with probability at most $t \times \big(\gamma(c'+1)\log n/m^\beta\big)^w$. Recall $t = m^\delta \ell$ and let $\ell = m^\alpha$. Fix w to be a constant so that $\beta w > \delta + \alpha$. Hence, with probability $(1 - o(1))$, every $C_i^{\eta\nu}$ has a decision tree of height less than w. Thus, C^ρ depends on at most $m^\delta \ell 2^w = O(m^\delta \ell)$ free variables.

We now show that this implies that almost all leaf-node labels have at least one free variable on which C^ρ does not depend. First note that the expected number of free variables in the label of any leaf-node of T_1^m is $pm/2 = m^{1-\beta}p_1/2$. Applying the Chernoff and union bound, with probability at least $1 - 2\ell \cdot \exp\big(-O(m^{1-\beta}p_1)\big)$, each leaf node label of T_1^m has at least $m^{1-\beta}p_1/2$ free variables. Call a depth-1 label bad, if C^ρ depends on every variable of the label. However, we observed that C^ρ depends on at most $m^\delta \ell 2^w = O(m^\delta \ell)$ free variables in total. Hence, the number of bad labels, denoted by N, is at most $O\big(m^{1-\beta_0}\ell/(p_1 m^{1-\beta})\big) = O\big(\frac{\ell}{m^\epsilon p_1}\big)$, where $\epsilon = \beta_0 - \beta > 0$. Since $p_1 = 1/O\big((\log n)^d\big)$, for constant d, only $\frac{1}{m^\epsilon p_1} = o(1)$ fraction of the ℓ depth-1 labels are bad. Thus, in each of at least $\ell(1 - o(1))$ many depth-1 labels, there exist a free variable on which C^ρ does not depend.

We now use some simple properties of our pointer function h^{PAR} to finish the argument. Recall that there are $\log \ell$ blocks of inputs to h^{PAR}, each of size $m/(\log \ell)$. Hence, again using the Chernoff and union bound, with probability $1 - o(1)$, each block has at least $m^{1-\beta}p_1/(2\log \ell)$ free variables. Hence, if we set the free variables of the label of the root note randomly, the root node pointer of T_1^m points to any fixed child node with equal probability, i.e. probability exactly $1/\ell$. In particular, with probability $1 - o(1)$, the pointer function at the root node evaluates to a child node whose label is not bad. This label, by definition, contains a variable on which the restricted circuit of Alice $C^{\rho\nu}$ does not depend. Hence, the output of Bob is correct only half the time for this case.

It is easy to verify that the above argument shows that the uniform distribution on the set of inputs of the tree function is hard. This is because our argument used random restrictions in

which conditioned on the fact that an input bit is fixed, it is fixed with equal probability to 1 or 0. Thus, we conclude that any deterministic 1-round compression game of cost $O\big(m^{\delta}\ell\big)$, gives the right answer on at most $1/2 + o(1)$ fraction of the inputs, if $\delta < 1$. □

4.2 r vs r+1 rounds

Here, we extend the previous argument above to arbitrary constant number of rounds:

Proof:[of Theorem 4.4] Note that we proved the base case of this for $r = 2$ in Lemma 4.5. We prove the general case by an inductive argument employing a round elimination technique. Let Σ be a $(r-1)$-round deterministic game of cost $O(m^{\delta}\ell)$ bits purportedly solving $TF_{r-1}^{m,\ell}\big(h^{\mathrm{PAR}}, Parity\big)$ of cost $O(m^{\delta}\ell)$ bits. We derive, by round elimination, a family of $(r-2)$-round games $\Sigma_1, \ldots, \Sigma_k$ for solving $TF_{r-2}^{m^{\theta},\ell}\big(h^{\mathrm{PAR}}, Parity\big)$, each of cost $O(m^{\delta}\ell)$ bits, where $\delta < \theta < 1$ are constants and the following holds: let c and c_i be respectively the probability that Σ and Σ_i give the right answer. Then, our reduction ensures that $c = max\{c_1, \ldots, c_k\} + o(1)$. This would complete the proof invoking the Inductive hypothesis, i.e. each $c_i \leq 1/2 + o(1)$. Below we first give the general idea of our round elimination technique employing random restrictions.

Let the compressor circuit employed by Alice for the first round of the game be $C = C_1, \ldots, C_t$, where $t = O(m^{\delta}\ell)$ and $\delta = 1 - \beta_0$, with $\beta_0 > 0$. Let each C_i have size at most n^c and depth d, where c, d are constants. We first show that one can hit C with a random restriction ρ that leaves many free variables and yet achieves the following: there are many subtrees such that the restricted circuit C^{ρ} of Alice depends only on very *few* variables occuring in the subtree. This will allow us to conclude that for almost all assignments to the free variables of the root node of tree T_{r-1}^{m}, the remaining $r-2$ rounds of the game Σ is solving a height $r-1$ tree function induced by such a subtree rooted at a child node of the root. Forthwith more details.

Mimicking the argument in the proof of Lemma 4.5, we first hit C with two rounds of random restriction. Let the resulting restriction be ρ, where ρ is sampled from Π_n^p, with $p = 1/\big(2m^{\beta}(2\gamma c')^d(\log n)^{d-1}\big)$, $\beta < \beta_0$, $c' > c + 2$. Then, just as before, the probability that some C_i^{ρ} does not have a decision tree of height less than w, is at most $O\big(t(1/n^{c'-c} + (c'+1)^w/m^{\beta w})\big)$. Choosing w to be a constant such that $\beta w > \delta + \alpha$, where $m^{\alpha} = \ell$, makes this probability tiny. Hence, with probability $(1 - o(1))$, C^{ρ} depends on at most $m^{\delta}\ell 2^w = O\big(m^{1-\beta_0}\ell\big)$ variables. Fix $\beta < \beta' < \beta_0$. Call a subtree rooted at a child node of the root of T_{r-1}^{m} to be bad if C^{ρ} depends on more than $m^{1-\beta'}$ variables occuring in the labels of this subtree. The total number of bad subtrees is at most $O\big(\frac{m^{1-\beta_0}\ell}{m^{1-\beta'}}\big) = O\big(\ell/m^{\beta_0-\beta'}\big)$. Hence, with probability $1 - o(1)$, at least $N = \ell\big(1 - O(1/m^{\beta_0-\beta'})\big)$ subtrees are not bad.

We next show how to derive our desired family of $r-2$ round compression games for tree functions induced by the non-bad subtrees. Recall there are $O(\ell^r)$ nodes in T_{r-1}^{m}. Further, the label of each non-leaf node is divided into blocks of size $m/(\log \ell)$. Hence, by Chernoff and union bounds, no block in a label of a node has less than $m^{1-\beta}/O((\log \ell)(\log n)^d)$ variables free in ρ with probability at least $1 - O(\ell^r \log \ell) \times \exp\big(-\Omega(m^{1-\beta}/((\log \ell)(\log n)^d))\big)$. Consider any assignment a of all free variables in labels of all non-root nodes of T_{r-1}^{m} that occur in bad subtrees rooted at level 1. Further, let b be an assignment to variables in the non-bad subtrees on which C^{ρ} depends. Note that b still leaves $O\big(m^{1-\beta}/((\log \ell)(\log n)^d)\big)$ variable free in each block of a label in a non-bad subtree. Now consider a random assignment x to the free variables of the root label of T_{r-1}^{m}. With probability 1, C^{ρ} is fixed and so the compressor circuit of Alice sends a fixed message. Bob's response also gets completely determined independent of how the yet unfixed variables belonging

to the non-bad subtrees rooted at level 1 are assigned. However, h^{PAR} evaluates to any given child node of the root with equal probability, i.e. probability $1/\ell$. Hence, we conclude that with probability $\big(1 - 1/m^{\beta_0 - \beta'}\big)$, $h_\rho^{\mathrm{PAR}}\big(x\big)$ activates a non-bad subtree of height $(r-1)$. Thus, with probability $(1 - o(1))$ the remaining rounds of the game yield a $(r-2)$-round compression of $TF_{r-1}^{O(m^\theta),\ell}\big(h, Parity\big)$ to $O(m^\delta \ell)$ many bits, for $\theta = 1 - \beta$. Note that by our setting of parameters, $\theta > \delta$. This completes the induction. □

Proof:[of Theorem 4.1] The proof of the upper bound follows easily using a straightforward deterministic protocol. The proof of the lower bound for $(r-1)$-rounds follows from Theorem 4.4 by appealing to the standard arument of showing that probabilistic games (circuits) imply deterministic games (non-uniform circuits) of the same cost that give the right answer on a fraction of the inputs that is bounded away from 1/2. □

5 Beyond correlation

Most of the techniques presented so far in this work for proving incompressibility, rely on methods that yield quite strong upper bounds on correlation.

Here, we take up one of the lowest complexity classes for which strong bounds on correlation are not known. Specifically, we consider the class of AC^0 circuits augmented with MOD_p gates, denoted by $\mathrm{ACC}^0[p]$, where p is an odd prime. The classical result of Smolensky [Smo87] yields that functions computed by such circuits of polynomial size and constant depth have correlation $O(1/\sqrt{n})$ with the parity function. This is a weak bound which cannot be used to prove incompressibility using the connection with correlation described in Section 3 of this work. In fact, to the best of our knowledge, no non-trivial lower bound was known for even the 1-round compressibility of Parity by such circuits before our work. Our main result in this sectiob provides such a lower bound. We make use of the following two results from the classical work of Razborov and Smolensky.

Theorem 5.1 (Razborov[Raz87],Smolensky[Smo87]) *Let f be any boolean function computed by an $ACC^0[p]$ circuit of constant depth and poly size. Then, there exists a $MOD-p$ polynomial P of degree $O(\log n)^{O(1)}$ that approximates f well, i.e. $\Pr_x\big[f(x) \neq P(x)\big] = O\big(1/2^{(\log n)^{O(1)}}\big)$.*

The above is complemented by the following inapproximability result:

Theorem 5.2 (Smolensky) *Let p be an odd prime and let P be a $MOD-p$ polynomial of $o(\sqrt{n})$ degree. Then, $\Pr_x\big[PARITY(x) \neq P(x)\big] \geq 1/2 - \Omega\big(1/\sqrt{n}\big)$.*

Combining the two above theorems, we show the following:

Theorem 5.3 (restatement of Theorem 1.3) *Let p be a fixed odd prime. The cost of any 1-round randomized $ACC^0[p]\big(poly(n)\big)$-compression game solving Parity is $\Omega\big(\sqrt{n}/(\log n)^{O(1)}\big)$.*

Proof: Let $C = C_1, \ldots, C_t$ be the 1-round compressor, where each C_i is an $\mathrm{ACC}^0[p]\big(\mathrm{poly}(n)\big)$ circuit. Using Theorem 5.1, we obtain polynomials $P_1, \ldots, P_t$, such that for each i, $\Pr_x\big[C_i(x) \neq P_i(x)\big] = O\big(1/2^{(\log n)^{O(1)}}\big)$ and degree of P_i is $O\big((\log n)^{O(1)}\big)$.

The first key observation is that the indicator function for the set of inputs that lead the compressor to output a fixed message has a low degree polynomial approximator. More precisely,

let a be any message and let $X_a \equiv \{x \in \{0,1\}^n \mid C(x) = a\}$. We construct a polynomial that approximates the indicator for X_a, denoted by 1_{X_a}, as follows: for each $i \leq t$, define polynomial $Q_i^a(x)$ to be $1 - P_i(x)$ if $a_i = 0$, else define it just to be $P_i(x)$. Then, it is easily verified that the following

$$1_{X_a}(x) = \prod_{i=1}^{t} Q_i^a(x).$$

holds for all x on which each $P_i(x) = C_i(x)$.

Let $A \subseteq \{0,1\}^t$, be the subset of messages for which the Solver outputs 1. Define,

$$Q(x) = \sum_{a \in A} \prod_{i=1}^{t} Q_i^a(x).$$

Thus, $Q(x) = Parity(x)$ holds for each x such that $P_i(x) = C_i(x)$ for all $i \leq t$ and the Solver gave the right answer on x in the compression game. Hence, $\Pr_x \left[Q(x) \neq Parity(x)\right] \leq \epsilon + t/\text{qpoly}(n)$, where ϵ is the error probability of the compression game. As error probability can be assumed to be $1/3$, Parity is approximated by $Q(x)$ on $2/3 - o(1)$ fraction of the inputs. However, the degree of $Q(x)$ is just $t(\log n)^{O(1)}$. If $t = \sqrt{n}/\left(\log n\right)^{\omega(1)}$, then we derive a contradiction invoking Theorem 5.2. This completes the argument. □

6 Open Problems

In general, one would like to understand better the connection between correlation and compression. While we showed tight lower bounds for AC^0-compression games using recent strong bounds on correlation between Parity and polynomial size AC^0 circuits, there are functions and circuit classes for which such bounds on correlation do not exist. One can, in principle, still hope to prove tight bounds in many such cases. For example for Majority, we obtained a tight compression bound by reduction from Parity. For separating the power of r rounds from $r-1$ rounds, we worked directly with random restrictions avoiding[4] a black-box usage of correlation bounds. However for $\text{ACC}^0[p]$ circuits, where proving strong correlation bounds is a major open problem, we could only show $\Omega\big(\sqrt{n}/(\log n)^{O(1)}\big)$ bounds on the 1-round compression. It would be interesting to tighten this bound. More so, as widely conjectured correlation bounds for $\text{ACC}^0[p]$ imply the imcompressibility of the Parity function by such circuits when p is an odd prime.

It would also be interesting to show tighter separations between r-round and $r-1$-round games. One natural approach, which has been used in the traditional communication complexity setting, is to use information cost arguments. It doesn't seem easy to apply such arguments directly in our setting - for example, in the AC^0-compression game for Parity, Alice can easily "randomize" the input by using a random self-reduction, thus communicating only 1 bit of information about the input. It's still possible that computational analogues of information cost could be used to make our arguments cleaner and to get tighter results.

[4]Note that our formulation of the connection between compression and correlation is insensitive to the number of rounds in the compression game.

Acknowledgements

The authors thank Jakob Nordström for inviting them to KTH Stockholm, where this work began. The second author would like to thank Yuval Ishai for suggesting it might be interesting to study compression games, and for preliminary discussions.

References

[AB09] S. Arora and B. Barak. *Complexity Theory: A Modern Approach*. Cambridge University Press, Cambridge, 2009.

[Ajt83] Miklos Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[BDFH08] Hans Bodlaender, Rod Downey, Mike Fellows, and Danny Hermelin. On problems without polynomial kernels. In *Proceedings of 35th International Colloquium on Automata, Languages and Programming*, pages 563–574, 2008.

[BH08] Harry Buhrman and John Hitchcock. NP-complete sets are exponentially dense unless NP $\subseteq$ co-NP/poly. In *Proceedings of 23rd Annual IEEE Conference on Computational Complexity*, pages 1–7, 2008.

[Cai86] Jin-Yi Cai. With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy. In *Proceeding of the 18th Annual ACM Symposium on Theory of Computing (STOC)*, volume 38, pages 21–29, 1986.

[CW04] Jin-Yi Cai and Osamu Watanabe. On proving circuit lower bounds against the polynomial-time hierarchy. *SIAM Journal on Computing*, 33(4):984–1009, 2004.

[CW09] Arkadev Chattopadhyay and Avi Wigderson. Linear systems over composite moduli. In *Proceedings of 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 43–52, 2009.

[DGS87] Pavol Duris, Zvi Galil, and Georg Schnitger. Lower bounds on communication complexity. *Information and Computation*, 73(1):1–22, 1987.

[DI06] Bella Dubrov and Yuval Ishai. On the randomness complexity of efficient sampling. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, pages 711–720, 2006.

[DP09] Devdatt Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[DvM10] Holger Dell and Dieter van Melkebeek. Satisfiability allows no non-trivial sparsification unless the polynomial hierarchy collapses. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 251–260, 2010.

[FRR+10] Sebastian Faust, Tal Rabin, Leonid Reyzin, Eran Tromer, and Vinod Vaikuntanathan. Protecting circuits from leakage: the computationally-bounded and noisy cases. In *Proceedings of EUROCRYPT*, pages 135–156, 2010.

[FS11a] Lance Fortnow and Rahul Santhanam. Infeasibility of instance compression and succinct PCPs for NP. *Journal of Computer and System Sciences*, 77(1):91–106, 2011.

[FS11b] Lance Fortnow and Rahul Santhanam. Robust simulations and significant separations. In *Proceedings of the 38th International Colloquium on Automata, Languages and Programming*, pages 569–580, 2011.

[FSS84] Merrick Furst, James Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, April 1984.

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, pages 6–20, 1986.

[Hås87] Johan Håstad. *Computational limitations of small depth circuits*. PhD thesis, MIT Press, 1987.

[HIK+07] Prahladh Harsha, Yuval Ishai, Joe Kilian, Kobi Nissim, and Srinivasan Venkatesh. Communication vs computation. *Computational Complexity*, 1(16):1–33, 2007.

[HMP+93] Andras Hajnal, Wolfgang Maass, Pavel Pudlak, Mario Szegedy, and Gyorgy Turan. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46(2):129–154, 1993.

[HN10] Danny Harnik and Moni Naor. On the compressibility of NP instances and cryptographic applications. *SIAM Journal on Computing*, 39(5):1667–1713, 2010.

[IMP12] Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC0. In *Proceedings of Symposium on Discrete Algorithms*, page To appear, 2012.

[Ish11] Yuval Ishai. Personal communication, 2011.

[KKL88] Jeff Kahn, Gil Kalai, and Nati Linial. The influence of variables on boolean functions. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[NW93] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM Journal on Computing*, 22(1):211–219, 1993.

[PS84] Christos Papadimitriou and Michael Sipser. Communication complexity. *Journal of Computer and System Sciences*, 28(2):260–269, 1984.

[Raz87] Alexander Razborov. Lower bounds on the size of bounded-depth networks over the complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the 19th Annual Symposium on Theory of Computing*, pages 77–82, 1987.

[Yao86] Andrew C.C. Yao. Separating the polynomial hierarchy by oracles: Part I. In *Proceedings of 26th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1–10, 1986.